



Citation, Citation, Citation: Bibliometrics, the Web and the Social Sciences and Humanities

Christine Kosmopoulos, Denise Pumain

► To cite this version:

Christine Kosmopoulos, Denise Pumain. Citation, Citation, Citation: Bibliometrics, the Web and the Social Sciences and Humanities. *Cybergeo: Revue européenne de géographie / European journal of geography*, 2007, 411, pp.13. halshs-00198734

HAL Id: halshs-00198734

<https://shs.hal.science/halshs-00198734>

Submitted on 17 Dec 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Citation, Citation, Citation: Bibliometrics, the Web and the Social Sciences and Humanities

Christine Kosmopoulos, CNRS-UMR Géographie-cités

Denise Pumain, Université Paris I, Institut Universitaire de France

pumain@parisgeo.cnrs.fr

Paru dans *Cybergeo, European Journal of Geography*, 12/12/2007, 411, 13 p.

Abstract

The use of digital resources and the affirmation of research assessment exercises throw a new light on the issue of bibliometrics. The paper reviews the main data bases and indicators in use. It demonstrates that these instruments give biased information about the scientific output of research in Social Sciences and Humanities. Emerging publishing and editing strategies on the web are analysed. The paper supports open access solutions and sharing resources policies for the social science's literature.

Key-words: bibliometrics, research assessment, scientific publication, web, Scopus, Google Scholar, Web of Science, social sciences and humanities

Résumé

La numérisation des ressources documentaires et la généralisation des procédures de l'évaluation scientifique renouvellent la question des usages de la bibliométrie. L'article recense les principaux outils de référence et les indicateurs en vogue. Il démontre que ces outils ne restituent que très imparfaitement la réalité de la production de recherche en sciences humaines et sociales. Il analyse les stratégies de publication émergentes sur internet et préconise des solutions pour un accès libre et partagé à l'information scientifique.

bibliométrie, publication scientifique, évaluation scientifique, Internet, Scopus, Google Scholar, Web of Science, sciences humaines et sociales

1 Introduction

In every area of social activity, systems of assessment are showing a tendency to provide further clarification of their objectives and methods, and even a greater transparency. Scientific production, whose functioning has depended for a long time on methods of evaluation by peers, is not exempt from this tendency. For several decades, certain disciplines have been setting in place a number of quantitative tools intended to measure the quality of the work of researchers and journals by making it possible to compare them by means of numbered indicators. For ten years, it seems, the 'digital' revolution, which has entailed not only an explosion in the number of publications accessible in digital form, but also a profound transformation in the methods of access to scientific information, has had to accelerate the use of bibliometric methods of

evaluation, even in the area of the human and social sciences that had previously remained reticent about this type of approach.

We can trace the origins of the practice of bibliometrics to the *Science Citation Index*, which at the outset was attempting to satisfy a highly specialized requirement for documentation, especially in the realm of the ‘hard’ sciences, a requirement that the professionals were not in a position to satisfy with their traditional instruments. This tool applied strictly to the analysis of scientific articles gradually evolved into bibliometrics. But with the social sciences, not only the objects of study, the research methods, and the relationship to the social context differ from those of the hard sciences, but also the modes of scientific communication. A report from Science-Metrix¹ on “The Use of Bibliometrics in the Social Sciences and the Humanities” illustrates the degree to which the modes of communication are much more heterogeneous in these disciplines than in the hard sciences (Archambault, Gagne, 2004). While assessment in the ‘hard’ sciences is based primarily on journals, the articles published in journals account for less than fifty percent of scientific communication, with books, chapters of books, monographs, talks given at colloquia, ‘grey market literature’ (informally published literature), and reports, accounting for more than half. For Hicks (1999), books have a much more significant impact in the social sciences, because they account for forty percent of the citations. And for that matter, these proportions, like research and assessment practices, vary among the social sciences along a wide spectrum, from economics through, for example, geography, to law.

The social sciences are confronted with diverging interests in the choice of a database of the bibliometric measurement type for reference citation, and in the degree of importance to give to these tools in the assessment of research. Private publishers attempt to impose their sources, and scientists who are already well established in these practices wish to disseminate their methods. Strategies adapted to the tools of the digital revolution are already very visibly in place, while an option of publication with free access is on the horizon. Will the social sciences manage to take advantage of this still uncertain context to evolve towards practices better adapted to their status?

2 Bibliometrics: databases and indexes

The social sciences as a whole are by profession wary of any form of bibliometrics. Dealing as they do with complex subjects, accustomed to the traps of ‘comparatism’ in history or between societies, and keenly conscious of the numerous influences from different contexts that must be taken into account, researchers in the social sciences are fully aware of the effects of reductionism that quantification can bring, in particular when it is used to examine a scientific property. The institutional organization of their knowledge and their competence through the internationalization of professional publications is much less advanced than is the case for the natural and life sciences. The diversity of local cultures, and the loss of meaning associated with translations from one language to another, are often adduced to challenge any attempt at bibliometric counting. Yet often as an easy solution and sometimes by calculation, a number of academic institutions, even in some non-Anglophone countries, have begun to use bibliometrics for classifying and recruiting researchers. The tools currently in use are still imperfect,

¹ A Canadian organization specializing in the evaluation of science with reference to Anglo-Saxon and Scandinavian countries.

however, and often unfair in the treatment of non-Francophone research, and of the social sciences in general. It is therefore important to set up an inventory of these tools so that a critical analysis of their use can be made. In several areas—at the level of databases, indexes and search engines—it is in fact possible to take steps to make sure that bibliometric tools do justice to the output of researchers in the field of the social sciences.

2.1 Bibliographical databases and the challenges they present

At present, three large databases of bibliographical information are available on the Internet. Set up by separate actors, they differ in the number and type of publications they register, which only partially overlap, and each gives only a very imperfect picture of scientific production in the field of the social sciences.

The first, and earliest, known today by the name of the *Web of Science (WOS)*, was originally set up by Eugène Garfield. As far back as 1955, in *Science*, E. Garfield made a new proposal for scientific communication: the creation of an index of citations from scientific publications. For the author, such an index satisfied the needs of scientists for specialized information, while the classical indexing system by librarians provided entries that were more for the generalist, and as a result were unsuitable. The objective as announced was highly pertinent: to make an inventory of the sources cited in an article, and also, the references correlated after its publication. Here it was a question of making known the network of scientific references and facilitating access to the information on a given theme, with the advantage for the hard sciences of having simultaneous access to the work of both applied research and fundamental research. This base, originally called '*Current Contents*', also had the stated mission to assist those making assessments to judge the pertinence of the bibliographical sources of the articles submitted to the journals (Garfield, 1964).

But the idea also satisfied two other objectives. The first of these, by means of an inventory and an evaluation of contents, was the provision of a form of monitoring of worldwide scientific literature. This is how, since its creation in 1964, the *Science Citation Index*, an extract of the journals indexed in the database of *Current Contents*, became established as a unique source for the citation of articles. The second objective, commercial in nature, offered a service supplying offprints of all the cited articles inventoried in *Current Contents*. This private company based in Philadelphia, known as the *International Scientific Institute (ISI)*, is now the property of the Thomson Group. At present it covers all the indexed titles for a period of sixty years, while also preparing a project for numbering the entire collection of documents in existence since the beginning of the twentieth century.

Recently, in 2004, a second database, *Scopus*, was set up by the publishing group Elsevier to compete with Thompson Scientific. The ambition of the commercial editor is to make it possible to access the whole of the world's institutional scientific literature. Unlike *WOS*, *Scopus* has no editorial board, but it has been indexing articles from reviews since 1996, according to a procedure that is very like that of *WOS*. *Scopus* indexes the titles of journals that are submitted as soon as they have been labeled by an institution. Today, in both cases, there is a charge for access to these private databases, which is negotiated (at top price) with the management of the research organization and university libraries.

The third source, *Google Scholar*, launched at the end of 2004, was not strictly speaking a bibliographical database in its initial conception; it is a search engine specializing in scientific literature. The database that grew out of it is the property of the group that launched the *Google* search engine in 1998. The indexed documents originate with scientific editors, scholarly societies, collections from open archives, universities and other research organizations. *Google Scholar*, like *Google*, uses a robot tool that proceeds to explore, classify and index the contents of a proposed site. For France, *Google Scholar* indexes the archives deposited on HAL (CCSD-CNRS)², a large part of the INIST's document collection (13 out of 20 million biographical notes), and the publications of the Revues.org³ portal according to the OAL protocol.

Ann Will Harzing, an Australian academic specializing in information technology developed independently a tool for text treatment and citation analysis based on *Google Scholar*, with the exuberant title 'Publish or Perish'(2007)⁴. Unlike the two databases previously mentioned, *Google Scholar* and *Publish or Perish* provide free access.

2.2 Indexes: evaluation tools

Computer programs for extracting information, which make it possible at one and the same time to index these publications in natural language and to calculate rapidly the number of times they are cited, are in fact the key to the growing success of these databases for establishing bibliometric tools. Their performance goes well beyond what the former classification systems of librarians could do, as natural languages are increasingly replaced by thesaurus classics or lists of key-words, with indexing now carried out not only according to the name of the author, the title of the journal and the book, the date of publication and editorial references, as well as the title of the article or the work, but also throughout the published text (*full text indexation*).

Data mining programs which, like *Google Scholar*, explore this gigantic body of information, sometimes only count the number of citations of an article found in the bibliographies of other articles or books. This is what the 'citation index' of the ISI offered in the beginning. But more sophisticated calculation tools are usually integrated into these research programs, even making a certain degree of analysis possible. The tools offered are based on recent studies analyzing the explosion of scientific observations in bibliometry⁵, exploration tools and meaning indicators having advanced together significantly over the last ten years.

The indexes establish the value of the journals, authors and researchers, based as they are on a simple preconception: that the number of times an article or work is cited is an indication of its usefulness and of its scientific value. The point is obviously open to discussion, as a large number of citations can also reflect controversy, refutations, or simply the effect of a fad, but on average it can be considered a positive indication of the reputation of a work. The number of citations per author makes it possible to establish the *impact coefficient* of a journal, providing evidence of its audience and of its capacity to

² <http://ccsd.cnrs.fr>

³ <http://revues.org>

⁴ We are grateful to Alain Peyraube for bringing this reference to our attention: <http://www.harzing.com/>

⁵ Manuel Durand-Barthez, *Citations et Facteurs d'impact: quel avenir pour l'évaluation?* ('Citations and Impact Factors: What Future for Evaluation?') online on the URFIST website: http://urfistinfo.blogs.com/urfist_info/2006/04/citations_et_fa.html

choose the best authors. Several specialist journals, such as *Scientometrics*, *Infometrics*, and *The Journal of Information Science* have been created to handle questions related to efforts to measure the value of scientific publications.

Already in 1935, a 'law' established by Bradford showed that reputation value is very unequally distributed, with 150 scientific journals out of several thousand accounting for half of the total number of all the citations. When the articles and books of an author are listed in a decreasing order according to the number of citations, the series usually begins with a relatively high number, from several dozen, even hundreds or thousands of citations for a reference that is often cited, but this number then diminishes quite rapidly for later writings by the same author. The series showing the number of citations is more like a decreasing geometric (or exponential) progression than a linear (arithmetical) diminution of the number of citations. Therefore, because the general form of the statistical distribution of the number of citations by article or book for the same author is not symmetrical, but on the contrary very dissymmetrical, with many articles receiving no citations or very few, alongside several that receive many, the number that can best represent this distribution is not the average number of citations per article, which would characterize the normal distribution of the number of citations, but a number taking into account the dissymmetrical form of this distribution. The physicist Hirsch (2005) proposed designating the number h as equal to the rank of the article that has received at least as many citations as the numerical value of its rank. This number in some way materializes the intersection of the curve representing the inverse geometric progression of the number of citations with the ascending arithmetical (linear) progression of the rank of the articles. The position of this number is representative of the scale of importance of the particular series of the number of citations. Mathematically, this number is close to the number of the logarithm representing the quantity of citations—but this last measurement would be less directly meaningful for those who use the index.

Although a relatively recent creation (2005), this index was considered sufficiently robust and explicit to be included the following year in the bibliometric calculations disseminated by Thomson/ISI on the *Web of Science*.

2.3 Insufficient coverage of the Social Sciences and Humanities (SSH)

Each of the three existing large databases of citations of scientific literature has substantial gaps, which make it impossible to establish a pertinent assessment of researchers or journals (*the impact factor*) by means of bibliometric tools.

Although considered by some to be the standard citation database, the *Web of Science* has been criticized for many years for its bias against the SSH. It inventories some 8,700 international journals (there are an estimated 20,000 scientific journals in the world), but only 1,000 for the SSH, almost all of which are of Anglo-Saxon origin. By way of example, the SSH journals supported by the CNRS do not appear in the *WOS*. In 2004, the CNRS tried to negotiate improvements with Thomson/ISI that would take into account specific European needs and particularities, but no progress has been made. A report by Philippe Jeannin on the evaluation of research in the SSH submitted in 2003 to the Ministry of Research and New Technologies confirms that French journals in these fields are not covered.

With a view to providing a better analysis of European scientific production in the *WOS*, the *European Scientific Foundation (ESF)* launched a program to evaluate journals

in the SSH, publishing in June 2007 an initial list called the *European Reference Index for the Humanities (ERIH)*⁶, which classes journals by rank (A, B, C). This list does not include certain research areas in the social sciences such as geography, which will become the subject of another classification. The ESF provides rulings concerning the integration of journals in this reference list, which explains why not all the journals supported by the CNRS figure in it. But if the ERIH does not have the vocation to be a bibliometric tool, and is not offered as an alternative to *WOS*, the *Observateur des Sciences et des Techniques (OST)*, which has the specific mission to conceive and produce indicators about research and development, has announced that it will be working on impact indexes based on the journals inventoried by the *ESF*⁷.

The *Scopus* database (Elsevier) could be an alternative to the monopoly of Thomson. In fact, *Scopus* indexes around 17,000 titles, of which 2,850 are in the SSH, that is, twice as many as the *WOS*, and is not limited to Anglo-Saxon journals. The geographical distribution of the titles is 25% for the United Kingdom, (4,157 journals), only 25% for the rest of Europe/Middle-East/Africa, 37% for North America, 12% for Asia/Pacific, and 1% for South America. There is therefore much broader coverage than that offered by the *WOS*. Furthermore, the inquiries made to this *Scopus* base associated with the search engine Scirus yield a much wider range of results, because *Scopus* also includes other documents than the articles in the journals. The list of the sources of Scirus is indicated on the site. Unfortunately, the period covered by *Scopus* remains very limited (11 years). A comparative study of the *WOS* and *Scopus* for ten disciplines in the SSH is currently underway under the direction of Christine Kosmopoulos.

When all is said and done, with both databases the coverage of the publications in the journals, both from the perspective of the journals covered and of the length of time covered, is so incomplete as to give a biased representation of scientific production in the SSH, and therefore, cannot provide solid results for a bibliometric evaluation.

Another advantage of *Google Scholar*, in addition to the free access it provides, is that it inventories all scientific literature without distinction: articles in journals, whether or not they have an editorial board, but also theses, books, extracts of books, reports, pre-prints, etc. Nevertheless, this specialized search engine presents other problems. Unlike the *WOS* or *Scopus*, it provides no information about the resources it uses. As Jean-Pierre Lardy indicates in his 2007 publication on the URFIST⁸ site, there is no list of commercial editors or the servers of the indexed archives, no information about the period covered, the volume, or even the countries involved. Certain known sources of *Google Scholar*, for example the bibliographical database *Francis* of the INIST, are not in conformity with bibliometric norms. There are also significant gaps in the coverage of publishers' archives. A test with reference to important figures in the sciences confirms this observation, as Peter Jacso demonstrates⁹. For this database to become an incontestable reference for citations, it would need a significant and systematic overhaul.

Because of the opacity of the sources and the incomplete coverage (not all of the publications necessarily appear in the same journal), the information extracted by the tool

⁶ <http://www.esf.org/research-areas/humanities/activities/research-infrastructure/faq-sheet/scope-initial-lists.html#c13190>

⁷ <http://www.obs-ost.fr/>

⁸ <http://urfist.univ-lyon1.fr/GoogleScholar.pdf>

⁹ <http://www2.hawaii.edu/~jacso/extra/gs/>

Google Scholar and analysed by Anne-Will Harzing on her website¹⁰ cannot yet guarantee a reliable evaluation. Yet Harzing's work does make a critical use of this database possible¹¹. The truth is that at this stage, the tool known as "*Google Scholar-Harzing*" gives a much better account of the various forms of communication in the SSH than the *WOS* or *Scopus*. Furthermore, its chief advantage is that it makes it possible to intervene in calculations about the index, in selecting the publications that are considered to be truly indicative of the scientific activity of an author, eliminating publications by authors with the same name, and duplications or references that are not really pertinent, and in proposing a whole variety of indexes that correct the h index. In this way, the g index (*Egghe's g-index*) is calculated on the same principle as index h, but gives more weight to frequently-cited articles. Another index (*Individual h-index*) corrects index h with the average number of authors per article, in order to make it possible to compare output between the disciplines in which the procedures for identifying the authors of an article are very different. A further feature of index h is its proper measuring according to the number of years of publishing activity of the author (*Age-weighted citation rate*), which entitles it to compare people who are in different phases of their scientific careers. Index h can also be calculated by assigning more weight to recent articles, giving a higher score to people who are still productive (*Contemporary h-index*).

To the extent that the *Google Scholar* database provides better coverage for publications in the human and social sciences, especially books, it would undoubtedly be useful to undertake for each discipline and sub-discipline a systematic test of its different measurements. For that matter, A. W. Harzing encourages this on her site. The value of the indexes can be considerably enhanced by a better knowledge of the quality of their presentation of scientific values habitually recognized by peers. Indeed, in spite of reticence about their use and the imperfection of the existing databases, it is highly probable that the expansion of online publications and the ease with which they can be consulted will soon lead to the adoption of citation indexes in institutional procedures for the recruitment and assessment of researchers. Instead of allowing themselves to be obliged to use the considerably biased instruments of the *Web of Science* that are presently used for a number of the natural and life sciences, it would be in the interests of researchers in the human and social sciences to appropriate tools developed from a database that is more open to their publication practices. Certainly all these quantified measurements cannot provide a complete substitute for the more qualitative evaluation, partly subjective, but much more subtle and sure, represented by the classic assessment by peers. Even in experimental disciplines, in which, it is generally admitted, the number of citations reflects quite well the quality of the production of a scientist, and in which the classification of journals is systematically taken into account in publication strategies, the tools of bibliometry are objects of recurring criticism. In fact, it is the social practice of scientific citation that warrants a critical examination.

3 Citations are a combat sport

Every scientific writer knows that the list of references he or she cites in an article in support of a given piece of work is partly arbitrary, and often full of gaps. In addition to the objective impossibility of knowing and citing the totality of useful references for a

¹⁰ http://www.harzing.com/resources.htm#/pop_gs.htm

¹¹ <http://www.harzing.com/>

given subject, the principal bias is doubtless related to discipline, each science deliberately ignoring advances made by the others, and all the more if that science occupies a higher position in the implicit hierarchy of accepted values (or rather, of the interest value) which would place economics, for example, on the top of the basket and the humanities on the bottom. But disciplinary orientations are not the only factor, and in professions in which the primary reward is the capital of reputation, citation is a strategic weapon. For several references of major influence, are there not any number of courtesy citations, deliberate omissions, internal favours exchanged within little fiefdoms that mutually ignore each other, coquetry in the form of 'exotic' citations (this is one of the facile explanations of the success of the *'French Theory'* in transatlantic or trans-Channel cultural studies), or even unfounded quarrels with an author of major reputation?

This is not the place even to sketch a sociology or a social psychology of the activity of scientific citation. We simply wish to point out the degree to which the habitual distortions of these practices, which are adopted by means of an 'objective' measure of reputation via the citations, are further amplified by being put online and by the globalization of access to references. New strategies adapted to these media have made their appearance. We know the manipulations of posting that made Google's initial research procedure possible, based on key words that needed only to be introduced in large quantities in hidden pages to cause a site to emerge among the first places. Since then, algorithms developed from semantic networks have been set up, making the means of acquiring visibility and being cited more complex, yet without blocking them. In actual fact, the proximity of information resources on the web, which researchers also use in carrying out their own bibliographical research, without necessarily limiting themselves to their usual scientific journals, reinforces the visibility of scientific 'products' when they are also often cited by other sources.

Among the possible ways of manipulating the bibliometric system, we could mention the one that consists of people who evaluate articles letting it be known that any article citing one of their own publications will be accepted (but perhaps this is only a rumour...). Another strategy consists in multiplying the possible areas of citation, by playing with uses of vocabulary that are on the outer edges of common usage or the expectations of the general public, something that occurs frequently in the human and social sciences. Part of the massive success of the author of the expression *'creative class'* (a 'new' sociological category that encompasses all types of creators, artists as well as entrepreneurs), attributing to them the responsibility of the dynamism of cities through innovation, can surely be explained as the result of a clever use of these combinations. The effect of being picked up by other media also increases the visibility of scientific articles that treat of 'subjects of society', like homosexuality, or again *gender studies*. The collusion of the media, voluntary or no, is frequent when researchers are working on subjects that sometimes overlap with topical events, whether risks, crises, criminality, terrorism, or even simple politics, especially around election time. The post-modern trend, which militated for the introduction of secular knowledge into the human and social sciences, by reproducing 'the words of the inhabitants', or in granting legitimacy to the knowledge content in discourse held by the people under observation, also contributes to a blurring of the frontiers, when the links and key words on the web point as much to different blogs as to articles online.

Will the speculative bubbles burst, or will they leave their mark? Will we see on the web an evolution of the biological type, by ‘natural selection’, with a period of settling that will allow the most ‘fit’ writings, whose quality enables them to feed the Internauts’ thirst for knowledge, to float to the top, or will we rather see a societal process established according to which ‘bad money drives out the good’? It would seem that on the web as well as elsewhere, ‘factoids’ are hard to get rid of, and that the correlation between media visibility and scientific quality is hardly increasing, justifying the pessimism expressed, for example, by Andrew Keen (2007). Will we one day have the technical means to detect authors of citations who are running around in circles, will we be able to offer correctives for semantic or paradigmatic ‘commercial’ networks, by means of the same efficient tools that are used for semantic analysis on the web? In the meantime, an interesting guarantee of quality is now being offered through the introduction of new practices in online production, monitored by peer evaluation. This means that scientific publishing will be able to free itself from commercial interests, thanks to ‘in-house’ validation of the expertise of researchers.

4 For a regime change in values

From now on, scientific communication will be using the *new technologies of Communication and Information* (NTCI), with the tangible result of a profound reform in methods of distributing scientific information. Most scientific work is digitized and put online either by researchers¹², or by commercial editors¹³, on the sites of open institutional archives¹⁴, on public sites¹⁵, or again on personal web pages. The NTCI are therefore modifying the behaviour of researchers, who now have the means of increasing the visibility of their work on a global scale, and of adding to their potential for being cited through online posting, but at one and the same time they have also brought about an important change in relations with commercial scientific publishers, as well as a change in work practices. In fact, the development of the NTCIs may well cause us to enter a new era, an era of sharing, that is, of *Free Access*. But this development is being blocked by the resistance of the financial actors, as evidenced in the recent creation by the *Association of American Publishers* of the anti-Open Access organization *PRISM* (*Partnership for Research Integrity in Science & Medicine*).

4.1 The sharing of resources

For several years, an arm-wrestling match has been going on, between commercial publishers who sell back to researchers/authors their own publications, having acquired the rights to them, and the international *Open Access* (OA) movement, made official by an appeal from Budapest in December 2001, which is demanding free access to all scientific documents (papers, theses, scientific articles, etc.)¹⁶. Nearly 400 universities and scientific institutes around the world are joining it.¹⁷ Numerous initiatives are moving

¹² Taking as an example the CNRS portal, revues.org

¹³ Cairn (a consortium of French publishers), Elsevier, Springer Verlag, etc.

¹⁴ For France, the CCSD of the CNRS

¹⁵ Persée (of The Ministry of National Education, of Higher Education and Research in France) is conducting a retrospective digitization of the journals in SHS.

¹⁶ <http://www.soros.org/openaccess>

¹⁷ <http://www.soros.org/openaccess/view.cfm>

in this direction, as for example the *Sherpa*¹⁸ project in England, whose mission is to develop pools of free-access digital academic documents, or *Driver*, a European infrastructure of support, with Open-Access digital reserves¹⁹. One can get an idea of available resources on OA by referring to *OpenDoar*, a worldwide repertoire of deposits in open archives²⁰, or the *Registry of Open Access Repositories* (ROAR)²¹.

The *Open Archive Initiative Protocol for Metadata Harvesting* (OAI-PMH)²² is a part of the movement of the *Open Archives* or *Open Access*. Its objective is to render the metadata on documents online interoperable, in order to permit the transfer of data between the different servers that respect this protocol. The principle is simple: the data are deposited on the server in a storage center and made accessible to all OAI harvesters who navigate (and reap!) on the web. Hundreds of sites now respect the OAI-PMH norms, for example, *Gallica* of the Bibliothèque Nationale de France, *Sudoc* (France), *Eprints* (England), the Library of Congress (U.S.A.), etc. The University of Michigan's scientific search engine *OAIster* searches in 714 databases of international data, and inventories more than 13 million references.

The idea of free access to scientific publications, which the Budapest appeal embodies, has been defended for many years by researchers; it took concrete form when the first server of the open archives deposit in Los Alamos *ArXiv*, was put on line by Paul Ginsparg in 1991. In 2001, the *Centre de la Communication Scientifique Directe* (CCSD)²³ of the CNRS in turn launched an open international multidisciplinary archive, based on free self-archiving. Respecting the OAI-PMH protocol, it facilitates exchanges between the large databases of international scientific data and the search engines²⁴.

Apart from the open archives that also include the sites of theses deposits, the number of electronic scientific journals in free access produced by researchers is increasing. 2,811 are inventoried in the DOAJ²⁵. In the field of the SSH, the oldest electronic journal, *Cybergeo: European Journal of Geography*²⁶, has been in existence since 1996. Recognized by the CNRS, it has an international editorial board, and publishes in all the European languages. Produced by academics and researchers independently of any commercial publisher, and offering free access, *Cybergeo* is listed in *Scopus*, and in *Google Scholar* via the portal of *Revues.org*, which it joined in 2007, and which applies the OAI-PMH protocol²⁷. Other initiatives are worth noting, such as *Hypergeo, encyclopédie plurilingue de géographie*²⁸ (*Hypergeo*, multilingual encyclopedia of

¹⁸ <http://www.sherpa.ac.uk/index.html>

¹⁹ <http://www.driver-support.eu/en/index.html>

²⁰ <http://www.opendoar.org/index.html>

²¹ <http://roar.eprints.org/>

²² <http://www.openarchives.org/>

²³ Centre of Direct Scientific Communication

²⁴ OAIster, Google Scholar, etc.

²⁵ <http://www.doaj.org/>

²⁶ <http://www.cybergeo.eu>

²⁷ *Cybergeo* has never been accepted by the evaluation committee of the *WOS*, even though the journal meets all the required criteria. The argument advanced to justify this is that the journal's impact factor in the inventory of the database of the *WOS* is nil. But as long as a journal is not integrated into the *WOS*, it cannot be cited there. The message is very clear. In reality, a scientific journal in SHS, and on top of that, a francophone one, which in addition offers free access, is of no more interest to Thomson than it was to ISI beforehand.

²⁸ <http://www.hypergeo.eu/>

geography), produced by universities for use by teachers and students, with the purpose of presenting the principal concepts and theories of geography.

4.2 Collective platforms with Free Access

With the impetus provided by the *Open Access Initiative*, therefore, sharing has many facets: open archives, publications offering free access, OAI-PMH protocol, but also collaborative tools, free computer programs (*Open Source*), and copyright arrangements of the *Creative Commons* type. Most of the platforms of digital resource management offered in 'open source', such as *DSpace*²⁹, that make it possible to archive, to index and to disseminate digital content, are compatible with the OAI-PMH protocol.

For collective work and the self-publication of documents, numerous applications are available in Free Access, among them *Google Document et Tableur* (Google Document and Spreadsheet), which facilitates the creation and introduction of documents, and makes it possible for several authors to work together. Certain laboratories are already using photograph-sharing sites like *Flickr* to manage and publish their collections. In the same vein there are the Wikis, which make it possible to create, modify and publish content in collaboration, the most well-known, of course, being Wikipedia, where anyone can interact with and add to the pages. Access to the visual display or the publishing feature of the pages may however be restricted by a password, as in the case of the scientific wikis organized by work groups on a given subject.

Other types of sharing platforms have emerged that give a clear indication of the change in direction, and of the revolution that is taking place in work methods. Online aggregating servers³⁰, for example, offer the possibility of sharing the syndicated feed³¹, if the subscriber wishes. Concretely, by subscribing to *Cybergeos*' RSS feed, it is possible to see posted on the site the number of subscribers to the feed list; clicking on the list causes the profiles to scroll through, and one can then access the interface of one of the selected profiles and read, use, or send back to one's own aggregating service the flow of information gathered by the person.

The same is true for 'Favorites' (Internet Explorer), or 'Bookmarks' (Firefox). Instead of keeping the page-markers in the local interface of one's browser, one can choose to collect them on a sharing platform like *del-icio-us* or *Google Reader*, with the advantage of being able to access them from any workstation online, and exchange data at the same time. *Connotea* also functions on this principle, but in addition offers a useful service for the management of references and articles online that was especially conceived with scientists in mind. As with the RSS feeds, users can choose to make their own bookmarks, classifications and references visible, with access to the complete text, and also access the texts of others.

5 Conclusion: Thinking bibliometrics in the context of NTCIs

The position of the social sciences in the systems of bibliometric analysis is not very secure, for reasons linked to differences in scientific practices, which are sometimes

²⁹ <http://www.dspace.org/>: a program for digital indexing and filing aimed at universities and research organizations.

³⁰ Examples: bloglines, lamooche, RSS-Feed, etc.

³¹ An RSS flow or syndication feed makes it possible to collect new elements placed online via an aggregating server, by replacing letters of information or warning systems via webmail.

attributed to a presumed ‘delay’ on the part of these disciplines in relation to what has been set in place in the disciplines known as the ‘hard’ sciences, but which probably also derive from other factors, in particular, cultural and linguistic diversity, and the more rapid historical development of research subjects. These subjects are also related to ‘society’, implying strong interaction between the object of scientific investigation and the social context in which the object is investigated (Latour, 1996). For that matter, some base their rejection of all bibliometrics in the human sciences on this supposedly insurmountable difference between them and the ‘historical sciences’ (the expression is Jean-Claude Passeron’s, 1991).

The development of collaborative tools, thanks to the second generation of the Internet (Web 2.0) and the movement of *Free Access*, and advances in the area of the internationalisation of the norms of exchange, offer new elements for reflection on bibliometrics in general, and open up perspectives for increasing the uses of bibliometrics in the human and social sciences. From now on, the technical means will make distribution possible, along with the sharing of scientific work at a low cost, and greater and greater access to resources, in particular via the OAI-PMH protocol or the RSS feed, with, as a result, an increase in the pool of citations, indispensable for arriving at solid analytical results. We can even imagine the formation of large new repositories or new databases that would combine existing databases (scientific search engines, open archives, libraries, etc.), as well as bibliographical references provided on the web pages of researchers. To this database of statistical information could even be added a directory based on peer opinion (Raaij, 2003).

An experiment conducted between 1999 and 2002 by the Universities of Southampton, Cornell and arXiv.org, *Open Citation Project*³², came to the conclusion that research on rows of citations and the links between the references is one of the examples of the OAI services, encouraging researchers to post their work on institutional sites. The HAL database launched by the CCSD of the CNRS is clearly a part of this project, and could serve as a bibliometric tool. Since the beginning of 2006, the CNRS has been conducting a large information campaign along these lines. The ESFRI (*European Strategy Forum on Research Infrastructures*) reflects on the new forms of evaluation of the SSH in the perspective of the 7th program-frame (7e Programme-Cadre) of European Union research³³. The European Foundation for Science undertook the indexing and validation of lists of journals for each area of the human and social sciences, with a classification based on their reputation.

It remains true that in the field of bibliometrics, as also for evaluation in general, trusting in a single indicator, however sophisticated, is an objective totally unsuited to what we know of the complexity of social systems, and that it would be advisable to set up not only batteries of indicators, but also multifarious methods for analyzing them and for preparing all decisions. Placing research in the network on a global scale thanks to electronic support and communication should cause the emergence of new forms of scientific evaluation, better harmonized, and of which the tools of bibliometry are only one aspect.

Archambault E., Gagne E., 2004 (dir), *L'utilisation de la bibliométrie dans les sciences sociales et les humanités* [‘The Use of Bibliometrics in the Social Sciences and

³² <http://opcit.eprints.org/>

³³ <http://cordis.europa.eu/esfri/> and http://cordis.europa.eu/fp7/home_fr.html

the Humanities’], Science-Metrix, Rapport final préparé pour le Conseil de recherches en sciences humaines du Canada (CRSH) [Final report prepared for the Canadian Research Council in Human Sciences (CRSH)] : http://www.science-metrix.com/fr/projets_2004_Bibliometrie_Sciences_Sociales.htm

Aubry C., Janik J. (dir), 2005, *Les archives ouvertes. Enjeux et pratiques*, Paris, ADBS éditions. [Open Archives. Challenges and practices]

Comité national d'évaluation de la recherche (CNER), 2006, Pour une meilleure évaluation de la recherche publique en sciences humaines et sociales, 72 p.: <http://www.cner.gouv.fr/cgi-bin/htsearch> [National Committee for the Assessment of Research (CNER), 2006, For better assessment of public research in the Human and Social Sciences, 72 pp.]

Garfield E., 1955, "Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas", *Science*, 122(3159), 108-111.

Garfield E., 1964, "Science Citation Index -- A New Dimension in Indexing", *Science*, 144(3619) 649-654.

Harzing A-W., 2007, "Reflections on the h-index". http://www.harzing.com/pop_index.htm.

Hicks D., 1999, "The difficulty of achieving full coverage of international social science literature and the bibliometric consequences", *Scientometrics*, 44(2), 193-215.

Keen, A. 2007. *The Cult of the Amateur: How the Democratization of the Digital World is Assaulting Our Economy, Our Culture, and Our Values*. Doubleday, Currency.

Jeannin Ph, 2003, *Revue-métrie de la recherche en SHS - Rapport synthétique et final de mission (1999-2003)*, Ministère délégué à la recherche et aux nouvelles technologies, France : [http://www.revues.org/cost/images/1/12/JEANNIN - Rapport_final2003a.pdf](http://www.revues.org/cost/images/1/12/JEANNIN_-_Rapport_final2003a.pdf) [Journalmetrics of research in the SSH – A Final Overall View of Mission (1999-2003), Ministry for Research and New Technologies], France

Latour B., 1996, *Petites Leçons de sociologie des sciences* [Little Lessons in the Sociology of Science], Paris, Le Seuil.

Passeron J.C. 1991, *Le raisonnement sociologique. L'espace non-popperien du raisonnement naturel* [Sociological Reasoning. The non-Popperian space of natural reasoning]. Paris, Nathan.

Van Raan, A.F.J., 2003, The use of bibliometric analysis in research performance assessment and monitoring of interdisciplinary scientific developments. *Technikfolgenabschätzung, Theorie und Praxis* 12(1), 20-29.